South Dakota Statewide Newspaper Digitization Project Planning Meeting
March 24, 2005

Participants

Julie Bolding, SD Dept. of Transportation, Pierre, SD
Greta Chapman, Director, Rapid City Public Library, Rapid City, SD
Lola Harens, Yankton Community Library, Yankton, SD
Jane Larson, Director, Vermillion Public Library, Vermillion, SD
Beverly Lewis, Director, Rawlins Municipal Library – Pierre (City Library), Pierre, SD
Dorothy Liegl, State Librarian, SD State Library, Pierre, SD
Miguel Penaranda, South Dakota State Library Digitization Dept., Pierre, SD
Renee Ponzio, Rapid City Public Library, Rapid City, SD
Marvene Riis, SD State Archives, Pierre, SD
Chelle Somsen, Director, SD State Archives, Pierre, SD
Colleen Smith, Director, Huron Public Library, Huron, SD
John Herbert, Marriott Library, U. of Utah
Liz Bishoff, OCLC
Tom Clareson, OCLC

Welcome & Introductions

Dorothy Liegl welcomed the participants to Pierre, and to the South Dakota Newspaper Digitization Project meeting. The meeting grew from an idea for a newspaper digitization project that originated with Colleen Smith, and the State Library is very supportive of the activity.

Purpose of the Meeting

Colleen Smith introduced the project, made possible because of funding support from the Daughters of the American Revolution and a matching grant from the Friends of the Huron Public Library. Colleen's vision for the newspaper digitization project is that it will include newspapers statewide. During the early planning stages, the Huron Public Library asked for assistance from OCLC, especially in creating collaborative projects. Two of the goals for this initial meeting are to give background information on newspaper digitization projects and to determine next steps in creating a statewide collaborative newspaper digitization project.

Meeting participants shared their interests, questions, and concerns in being part of a statewide newspaper digitization project.

Renee Ponzio: The Rapid City Public Library is interested in digitizing an entire run of papers, both historical and current.
Miguel Penaranda: The State Library Digitization Department would like to learn how to make digitized newspapers more widely accessible and available.
Beverly Lewis: From the viewpoint of patrons, reliability of access, easy searchability, and access to all issues are most important. While availability depends on funding, patrons want it all as soon as possible.
Jane Larson: Easy access and patron usability are key; the Vermillion Public Library would like to start with historic issues.
Dorothy Liegl: The State Library is looking at three different projects. The first project is from a historical perspective, and will involve materials prior to 1923. The second project addresses the need for access to current materials. The South Dakota Newspaper Association will digitize current materials from small newspapers, but not from the larger ones, although some are available from commercial sources. They need the *Rapid City Journal* and other large newspapers. The third project focus is on materials from 1923 forward.

Lola Harens: One important focus is for the index to be comprehensive. Genealogists looking for family members should be able to find information, such as obituaries, birth, and death records, which is important to them.
Julie Boldings: The Department of Transportation is interested in preserving the institutional memory that was created through the South Dakota Newspaper Project. Through that project, newspapers were cataloged and information on holdings was collected. The quality of the microfilm held at the State Archives was also analyzed; some of the microfilms are blurry. One decision that will need to be made is how to address low-quality microfilm. The National Digital Newspaper Project is digitizing from microfilm; this will be useful as a test to see if doing so is a viable approach. Because of the variety and quality of microfilm, it may be necessary in some cases to digitize from the original papers.
Chelle Somsen: The State Archives is interested to know what role the South Dakota State Archives will play in this project since they have the microfilm masters of newspapers. Microfilm is currently being used for sales, interlibrary loan, and research, and is a revenue source. The State Archives also charges for photocopies in their Reading Room. Determining how to generate future revenue is a big issue for the project.

Improving Access to South Dakota History Through Newspapers

John Herbert, of the University of Utah, spoke about the Utah Digital Newspapers Program.

The Marriott Library at the University of Utah scans historic Utah newspapers, makes them searchable, and loads them onto a website to create the Utah Digital Newspapers Program. The program URL is http://digitalnewspapers.org and it can be accessed from the Marriott Library's home page: http://www.lib.utah.edu. Utah Digital Newspapers is the first hit when the search term "digital newspapers" is entered into Google, Yahoo, and MSN.

Current holdings include 35 newspaper titles for a total of 253,000 pages, with publication dates ranging from 1858 through 1961. Twenty-two of the twenty-nine counties in Utah are represented. The statewide initiative expects to have twenty-seven counties represented by October 2005, with page growth predicted to reach 407,000 pages. With the digitization of the *Deseret News*, which began publication in 1850, the date range represented will be even broader.

Current website usage is up nine times from the usage in mid-2003. In June 2003, the site received 1,800 visits (an average of sixty per day); in February 2005, the site received 15,500 visits – an average of 553 per day. The site was featured on KSL-TV in October 2004.

The project began with an LSTA grant in 2002. This grant, totaling $93,000, was used to develop the digitization process, under the leadership of Kenning Arlitsch. In December 2002, the website was launched, with three rural weeklies on the site. 10,000 pages were digitized, representing thirty years of content from each of the newspapers. In 2003, a second LSTA grant of $278,000 enabled the project to add another 106,000 pages, representing eighteen new titles from ten counties. The grant also provided funding to launch a publicity campaign, including the creation of a paper published in *Serials Librarian.* Also in 2003, Utah Digital Newspapers was awarded a two-year IMLS grant in the amount of $1,000,000.

The project continued its growth curve in 2004 with the addition of eleven new titles and 115,000 pages. A third LSTA grant, for $99,000 was awarded; the project also received grass-roots support from five different counties. The project received an Award of Merit from the American Association for State and Local History (AASLH). A paper about the project was published in

Microform and Imaging Review. This article discusses technical specifications and is available on the UDNP website.

Significant finds include two newspapers – the *Richfield Reaper* and the *Murray Eagle.* The *Richfield Reaper* had a ten-year gap, from 1904-1914, that has never been microfilmed, and is now being digitized.

The next phase of the program includes focusing on distribution and aggregation. Holdings will be distributed across the state, and CONTENTdm will be used to allow users to search across all papers and go back to other collections to perform searches as well. This is possible because CONTENTdm has built this feature into its multi-site server. Brigham Young University will handle the *Deseret News,* the University of Southern Utah will handle the *Utah/Logan Journal,* and Southern Utah University will take care of the southern Utah newspapers. By summer 2005 all four locations (including the University of Utah) will be active. A training program is under development and will be ready by the end of 2005. In addition, the project was awarded one of six National Digital Newspaper Project (NDNP) test-bed grants. Finally, the University of Utah will host a newspaper conference sponsored by the International Federation of Library Associations Newspaper Group in 2006.

Users can find rich and varied resources at the Utah Digital Newspaper site, including some references to South Dakota. Some of the search phrases, and the number of hits, include:

- Butch Cassidy – 65 hits
- Titanic – 826 hits
- President McKinley – 1,776 hits
- Theodore Roosevelt – 2,638 hits
- Statehood – 2,478 hits
- Manifesto – 1,422 hits
- Earthquake – 5,946 hits
- South Dakota – 5,182 hits
- Black Hills – 753 hits
- Mount Rushmore – 15 hits

The UDNP launched a brief (10-12 question) user survey in 2004. User feedback yields some interesting statistics. 63% of visitors to the site visit at least monthly; 61% of the visitors are looking for genealogy or family history information; and 78% of the visitors find at least some of what they were looking for. 85% of the visitors rate the site as either "good" or "excellent;" 75% will share information about the site with others. And finally, 32% of the site traffic is from people who live outside of Utah. The improvement most often (by far) suggested is to add more content to the site.

The discussion shifted to how teachers might use the content found in historical newspapers. Colorado's digital newspaper project built in lessons for teachers that showed how they could use the web-based information and how to use historic newspapers. Newspapers' use of incorrect information and the pejorative terms found in old papers is helpful in literacy education. In a South Dakota example, Vermillion Democratic and Republican newspapers covered the same story differently. South Dakota will need to consider if this type of information will be included in the project.

John then returned to his presentation to discuss technical issues. Scanning was done at 400 dpi for 4-bit grayscale images. With I-Archives as their scanning vendor, UDNP gets three scans, at three different settings, per page (if needed) from the film. The quality of the photographic image depends on how it is displayed on the web page. Costs were estimated at $0.22 per page for scanning from film. Scanning from original newspapers was slightly higher at $0.30 per page. Image preparation included cropping, de-skewing and de-speckling. Full-page images (tiff files) are archived to LTO tapes. Article zoning and classification was done for news, ads, birth, death, and marriage records. Masthead information and article headlines are double-keyed and reconciled; headlines are accurate to almost 100% in their database. Optical Character Recognition (OCR) is generated on the whole page; it is uncorrected for text, but the headlines are, as previously mentioned, double-keyed. OCR standards include up to two words from two engines. The text is filtered through dictionaries, including English, place names and surnames. PDF files are created with embedded text in bi-tonal images for speed of download. Plans are in place to migrate to jpeg 2000. Image preparation costs for OCR are estimated at $1.27 per page. Database loading costs (DiMeMa indexes for CONTENTdm) are $0.15 per page.

Compression approaches vary; Utah is using the method which is easiest for users to download. All of the digitization and conversion work is outsourced to I-Archives; the University of Utah does not do scanning or OCR. I-Archives sends the OCR data to DiMeMa on DVDs via Federal Express. It is indexed for CONTENTdm, and then a different set of DVDs is generated.

Some of the website features include the ability to search across all papers from the main web page. This is the site's most powerful feature. In addition, the main page lets visitors quickly see which titles have been recently added, as well as titles to be added in the future. With 63% of the site traffic coming from repeat visitors, this is a useful feature. The main page also has a link to the web survey and a robust help link.

Each newspaper title has its own page on the site. Viewers can browse by issue date; search headlines or full text, and search or browse genealogical information, including births, weddings, and deaths. The site contains a County Map listing digitized newspapers from that county. This is the sixth most popular page.

Utah prefers to digitize from original papers because they can take a digital photograph of a page; this photo meets twenty-first century standards. In addition, OCR is cleaner on original photographs. In deciding to digitize from microfilm, or from original paper, the UDN notes the following pros and cons. Microfilm may be more readily available, and is cheaper and faster to scan. In addition, scanning can be done off-site, since microfilm is easily transportable. Microfilm is also NDNP compatible. However, the quality of the digital images depends on the quality of the film image, which can vary widely. The United States Newspaper Project (USNP) set microfilm standards in the 1980s; many images are decades old. Finally, according to UDN tests, search accuracy from microfilm images is 10% lower than from images digitized from original paper.

Scanning from original paper allows for the creation of a clean, new, digital image. The quality of the digital image is more easily controlled, and the search accuracy is higher than with images scanned from microfilm. However, it is harder to find originals of newspapers, and when found, these papers may need repairs prior to scanning. Although it is more difficult to find firms to scan original papers, and flatbed scanners are not a good option for scanning, the digitization is more likely to need to be done by a local source. It is more expensive to scan from original paper, and finally, the NDNP is not providing funding for scanning from originals.

Issues to be considered when launching a digital newspaper collection include locating and working with sources of funding. The National Digital Newspaper Project (NDNP) adds a big, long-term dimension to projects. Other issues include selecting a software platform, making decisions about hosting the website and the data (which includes looking at maintenance and sustainability costs). If you choose a distributed, rather than a centralized model, be aware that a distributed collection requires the creation of an aggregated index. Website design and decisions about whether to process the collection in-house or use outside service providers for scanning, OCR, and database loading are also part of the planning process.

Standards, including specifications for archival and web images, need to be set. The level of segmentation and classification needs to be determined. Content selection decisions will focus on which titles will be digitized, and in what order, and whether these titles exist in film or as originals. There are potential copyright issues to be considered as the selection process begins. It is best to have a strong project team in place, including a project manager, web developer, and a systems administrator. An Advisory Board, consisting of librarians, historians, and preservationists, among others, will be of value as the project unfolds. And finally, the special needs of the citizens who will be using the site need to be considered.

Options for Accessing Historic Newspapers

Liz Bishoff gave a presentation on options for accessing historic newspapers.

When planning a historical newspaper digitization project, there are many considerations. One of the most important early considerations for project collaborators is the development of a common vision. Each project needs to define its purpose, goals, audience, the environment in which it exists, and how it will be funded, both in the development phase, and looking forward, as it reaches maturity. Although these may have been articulated for the brick and mortar environment, they may change in the Internet environment, and free text searching opens a whole new world.

It is important to understand the source of the content, be it microfilm or original paper, and the challenges inherent with both sources. It is also important to determine who holds the master microfilm. Limitations, including copyright and intellectual property rights, need to be taken into account. And the project participants need to have a clear understanding of the opportunities, including asking themselves if there is a window of opportunity to meet an identified educational need, or a legislative or funding opportunity that can be leveraged to contribute to the project's success. In Colorado, where the history of library funding from the legislature is not good, the legislature would not fund newspaper digitization activities. However, the project has received tremendous community funding support, including a "Dollar-a-Page" campaign for digitizing local communities' newspapers. In South Dakota, money from gambling in Deadwood goes for historic preservation. It may be possible to use some of these funds to help support newspaper microfilm and digitization activities.

It is important to perform product comparisons. Existing microfilm may or may not be adequate for the project's purpose. If there is existing software, it is beneficial to document the strengths and limitations of that software regarding its ability to perform searches and retrieval of newspaper content. Comparisons of the benefits and costs of various services, and options (with benefits and costs) for partnering with commercial online subscription newspaper databases, are important. Estimates of the product requirements and costs to create the database also need to be done.

The California State Library newspaper digitization project (www.cpc.stanford.edu/cndp) found that microfilm was generally adequate for image capture. They also found that it was difficult to perform a product-by-product comparison, as most of the products examined required additional development. The solutions presented were different enough to make analysis difficult; the costs varied and were dependent upon the content management system and the provider.

The California State Library found that costs varied most when specialized search capabilities and sophisticated retrieval were offered. As levels of specialization and sophistication rose, costs also escalated for digitization and image processing. They found that if it is necessary to find a provider to host the service, a Request for Proposal (RFP) needs to be very specific. Most vendors did not understand the need for preserving digital content beyond providing security, back-up, and periodic copying. It is important to have a plan for migrating to new platforms, and to do constant checks to ensure the authenticity of files. The South Dakota State Library currently saves documents as pdf files stored on CDs or DVDs; Bishoff noted that at this time DVDs are not considered to be a preservation medium.

When evaluating products, there are some essential features to be examined. These include the product's search capabilities, including the abiltiy to search any words or an exact phrase, both for titles and full text. Searchability is very dependent upon the quality of the OCR and re-keying. Other considerations are the search date features, including start and end dates; the search results, and whether they provide the publication name, date, and a link to the article; the ability to locate the hit within the article for context; the navigational capabilities, including between pages, within a page, and the ease of motion; and the ability to browse among dates and identify holdings for each title. Usability considerations include features that are easy to understand, intuitive icon labels, minimal clicks to results, and the amount of real estate that is available for text display of the paper. The final essential feature is text legibility. It is important that the text be presented to facilitate reading, including offering unbroken images and good contrast. Because the user base for the South Dakota project is likely to include older users who are perhaps not as computer literate as younger audiences might be, it will be important that the site be as intuitive to use as possible.

Less important (but by no means unimportant) features identified by survey participants include the ability to search by article types, such as classified ads, birth, death, marriage, and advertising, The number of documents retrieved and the limit to the number of hits in response to the search were also features to be considered.

The least important features were the ability to limit by language, the document file size, and the thumbnail image.

Vendors provide various services, including:

- Hosted services that provide conversion and server hosting; these often offer limited customization. Heritage Microfilm is an example of this type of service.
- Software services offer content management systems, and include Greenstone, an open source software provider, and CONTENTdm.
- Software and conversion services: Olive Software is one provider.
- Conversion only is offered by vendors like I-Archive and Techbooks.
- Companies such as OCLC and Northern Micrographics offer digitization scanning services. OCLC has served as a "general contractor" on projects, putting together a team of various service providers for a project.

Most of the services are combinations, such as software and conversion, and frequently, multiple vendors may partner in response to an RFP. Be sure to ask if vendors re-sell your content to other users, and be sure the file they provide as back-up is created in compliance with national standards.

Digitization usually refers to scanning; conversion refers to OCR. Conversion can be done from microfilm, paper, or both. As a Request for Proposal (RFP) or a Request for Information (RFI) is created, be aware that it is possible to request a test for a small file at no cost. Tests of more than 50 pages are usually done for a fee. The test will generally be mounted on the vendor's site, but can be sent to the requesting institution for mounting.

It is important to tie quality control to standards, such as using TIFF images as master images. Quality control needs to be performed on scanning, metadata creation, and OCR keying, if appropriate. In addition to the quality control done in-house, vendors should have quality control processes built into the project, and should be able to articulate these processes. Quality control issues connected to digital preservation include checking master TIFF images on intake, creating a preservation action plan and reviewing data annually, and making plans for migration.

Regarding metadata and OCR, it is important to create a metadata record with a text description. This should meet existing standards and provide rich description. It can be combined with automated capture of data, such as the newspaper title, date, and page numbers. Subject classification can be used with local thesaurus terms, and MARC records or other indexing data can be imported. The metadata record becomes the master text file; easily converted to XML structured data, and to an approved WorldCat record. One problem that occurs with historic newspapers is the clash between 19th century vocabulary and 21st century users. For example, Denver, Colorado was not always named Denver. Project participants will need to think about how to deal with vocabulary issues for an audience that is key-word dependent. There are multiple approaches to dealing with this issue. Colorado is adding search terms to accommodate their users. The *Brooklyn Eagle* offers instructional information on vocabulary changes.

One decision point regarding OCR is deciding how users will be able to retrieve information. There are a number of different sections and items that make up a newspaper and a newspaper page, including headlines, articles, advertisements, illustrations, and photos. Retrieval choices include page-based retrieval or item-based; either may be appropriate for the audience.

The final issue is how to handle migration to new formats and platforms in the future. The platform or format selected at the beginning of the project will not be the last platform or format used – and you want to be sure you will not have to re-scan the collection. Therefore, make sure that the project begins with non-proprietary formats, such as TIFF and XML, and perform high-resolution scanning. The time spent on project planning will not be wasted time.

LC/NEH National Digital Newspaper Project

Tom Clareson gave a presentation on the Library of Congress/National Endowment for the Humanities National Digital Newspaper Project (LC/NEH NDNP). The NDNP has national implications for newspaper digitization.

The National Digital Newspaper Program, which was initially announced less than a year ago, is a natural outgrowth of the U.S. Newspaper Program (USNP) microfilming efforts. The NDNP will take the microfilm output from the successful USNP and provide much greater access through digital technology. Like the USNP, the NDNP wants to be as inclusive of all types of

news publications as possible. And, like the USNP, a lengthy, phased approach will allow for development of standards and good practices, replicable from state to state.

The NDNP has, as its mission and vision, goals of enhancing access to U.S. newspapers, using new technologies to further access to the USNP, the development of newspaper digitization standards and best practices, the phased development of the program in order to allow for research and scaled development. Through a geographically diverse approach, the NDNP will represent all U.S. communities.

There are a number of reasons behind the National Digital Newspaper Program which echo the justification for the USNP: enormous, widely-distributed collections and the huge amount of publications and pages; the fact that no single institution holds the "master collection" of newspapers – while the Library of Congress comes closest to "having everything," they want to ensure that access is provided to all; the wide-ranging user population for newspapers, including researchers, genealogists, and scholars; the importance of local and state-by-state selection to build a national program; the importance of full, comprehensive chronological coverage; and finally, building a leadership based on the USNP, utilizing the knowledge of institutions who managed the USNP for the National Digital Newspaper Program.

The key institutions driving the new National Digital Newspaper Program are the Library of Congress and the National Endowment for the Humanities Division of Preservation and Access. These two organizations, in their work on the U.S. Newspaper Program, developed a strong, standards-creating group of leaders. The USNP was a 20-year program that, in addition to creating standards for preservation microfilming, cataloging, and descriptive metadata, also developed a national cooperative network of experts.

The NEH also has experience with massively-sized projects such as the Brittle Books microfilming program. The Library of Congress actually tested newspaper digitization methods as part of digitizing the *Stars & Stripes* military newspaper for the American Memory Project.

When the twenty years of the NDNP activities have elapsed, here are some of the work products that will have been generated:

- Web-based access to
    - Over 30 million page images of historical newspapers, digitized primarily from microfilm, with full-text searchability.
    - A national directory of U.S. newspaper holdings, based on legacy data from USNP.
    - Historical information and context on newspaper publishing in the United States, printing technologies, and other topics.
- A depository of duplicate digitized microfilm at the Library of Congress.

Early information on the program has indicated that newspapers published between 1836 and 1922 will be included in the project as it continues.

The NDNP will use many partners. The National Endowment for the Humanities Division of Preservation and Access is providing program funding through its "We The People" initiative. The Library of Congress will aggregate, serve, and preserve the data. And the state awardees will select and convert the newspapers for the program. Just as with the USNP, LC and NEH will work with state-based groups to make the NDNP a success.

State groups in South Dakota and elsewhere will select newspapers to be digitized and will coordinate the conversion process (with NEH funding). The Library of Congress will serve the digitized newspapers via their web site.

Phase I of the NDNP is less than a year old. The program was announced at the American Library Association's Annual Meeting in June 2004. In July 2004, NEH Cooperative Guidelines were issued, and the Library of Congress developed the technical architecture. The deadline for the Phase I grant applications was October 1, 2004; 15 proposals were received. Those states receiving awards have just found out; a public announcement will be made in late March 2005. Grant recipients will meet as a group for the first time in May 2005. The second round of grant applications will be due in 2007, with applications available in September 2006 through the LC and NEH websites.

Fifteen states submitted proposals for this first round of funding. Six grants will be awarded. Projects with existing newspaper digitization experience and who have possession of master microfilm negatives were given preference. Grant awards are up to $500,000 each.

Each state has a goal of digitizing 100,000 pages by 2007, with dates ranging from 1900 – 1910. Analysis of the microfilm condition and other research will be included in this initial work, and the Library of Congress is creating some additional content.

Included in the National Digital Newspaper Project will be a new phase for the Newspaper Title Directory. The Newspaper Union List, created under the USNP and maintained by OCLC, will be reused along with CONSER (Cooperative Online Serials) records. This represents nearly 150,000 newspaper titles and 900,000 holdings records. The NDNP will provide searchable web access to all data collected by USNP, tied to digitized issues when available, and also to external newspaper web sites.

Early newspaper digitization efforts are finding gaps in USNP filming and coverage, which will help with completeness of runs in the future by identifying missing issues to be microfilmed and digitized.

While the searchability issue of page-level instead of article-level caused some concern from states which had already begun newspaper digitization, LC/NEH thought that offering full text with page-level access

- Preserves the integrity of text in context, primarily historical content
- Achieves reasonable search results with minimal metadata requirements
- Achieves economies of large-format, large-scale digitization, and
- Creates a substantial base of content for research and development on additional technologies and search strategies.

When the Library of Congress digitized *Stars & Stripes*, they found no significant difference in retrieval despite not offering article-level retrieval.

Technical specifications, which were included in the meeting participants' handouts, are the base-level requirements for each state project. To be "NDNP-compliant," however, some of the OCR specifications are proving difficult for vendors to meet.

After the initial test, or pilot phase, the National Digital Newspaper Program plans on adding new partners, awarding grants in all U.S. states and territories can digitize their papers.

It is expected that standards, technology, workflow, and production will improve as the future phases of the project move forward.

The best contact to find out more about the NDNP is Deborah Thomas, Digital Projects Coordinator for the NDNP, who has begun to make many presentations on the project. She can be reached via e-mail at deth@loc.gov. Additional funding information is available on the NEH website at www.neh.gov.

All this brings about some issues for consideration:

- Do South Dakota libraries and archives want to wait until 2007, or possibly later, to begin their newspaper digitization activities as part of the NDNP?
- What can be done now to get started?
- If South Dakota moves forward with projects prior to 2007,
    - Where will funding come from?
    - Will the project be NDNP-compliant?

John Herbert offered some observations that address some of the issues raised. JPEG 2000 will help with navigation of page-level segmented papers. The NDNP is not funding article-level segmentation; that has to be done locally. It is possible to pay locally to put these in a local database, and to do article-level segmentation that way, while still using NEH funding to get scanning and metadata done. The NEH grant covers 80% of project funding.

CONTENTdm does article-level segmenting; it can be changed to meet LC/NEH considerations. LC will support some legacy input of already-digitized materials. However, it is important to be sure that projects are NDNP-compliant.

At this point in the meeting, Tom Clareson gave a brief demonstration of a small digitized database of South Dakota newspapers available via CONTENTdm

Opportunities Analysis Discussion

Tom Clareson's presentation was followed by a discussion of the opportunities in newspaper digitization available to South Dakota.

Dorothy Liegl opened the discussion by asking whether the preference among the group would be to digitize all the newspapers in the state, or to focus on longer runs for certain papers.

Greta Chapman feels that the greatest interest is in scanning papers from places that are no longer in existence, where the newspaper is the only piece of history. Important periods in history was a focus as Colorado began their newspaper digitization project, but in the end they decided to digitize the first issue of the first paper.

Lola Harens asked where the database should be based. The South Dakota Library Network seems to be a logical choice.

Greta Chapman suggested looking to the Black Hills, gambling money, or the state legislators for funding. State LSTA money might also be a source of funding.

The group performed a SWOT (Strengths, Weaknesses, Opportunities, Threats) analysis, the results of which are listed below.

Strengths

- Microfilm exists in one place
- Committed partners in place
- Collaborative environment
- Technology exists
- Willing funders for collaboration
- Collaboration on SDLN as possible delivery mechanism/host
- Pride in history, especially in small communities
- Knowledge base about collections

Opportunities

- Disappearing communities; newspapers offer history
- Gambling money – Black Hills & Legislative
- Pick longer runs
- County-based approach
- Do planning – start-up projects to prepare for next round of NDNP
- Outreach to genealogical societies
- Global/24-7 reach
- People who have left can reconnect
- Online donation potential
- Preservation via new technology

Weaknesses

- Not a wealthy state
- Publicity/PR/Advocacy – must be on the road
    - Challenge: Fund some level of staff person
- Staff/Time
- Making project a priority
- Educating public on how to do research with newspapers
- Quality of film – potential concern
    - Degradation
    - Poor original filming

Threats

- Revenue concerns
- Deterioration of originals – newspapers and film
- Long-term commitment (challenge)
- Challenges
    - To get everyone at the table and reach acceptable compromises
        - Academic libraries
        - SD Humanities
        - Center for the Book
    - Multi-type nature of project

South Dakota Newspaper Activities to Date

Dorothy Liegl and Colleen Smith gave an update on recent activities, including the grants Colleen has received from the DAR and the Friends of the Huron Public Library. There are several events where the SDNP might be publicized, including the State History Conference in April 2005, the South Dakota Library Association meeting and the "History Conference in the Hills" in October 2005, and perhaps even at the South Dakota State Fair.

Liz Bishoff suggested that a website about the project be created, with a link to a CONTENTdm project. It will be important to set people's expectations as to the size of the project, which is best limited to 2,400 pages.

A rough estimate of the number of microfilm reels for the initial project, (alphabetically by newspaper title through P) is 4900 reels, but there is no estimate of the number of pages per reel. The Rapid City Public Library has a list to 1923 of reels. The group needs to identify who has microfilming capabilities, and add an idea of counties where paper is located. There are possibly 3.1 million pages.

Copyright and Other Issues

Greta Chapman asked the group where they see themselves as being with public domain newspapers. The project can be worked on from several fronts; with federal funding filling in pieces of state money or vice versa. People will want more than just newspapers in the public domain. Greta suggested a three-phase project:

1) Public Domain newspapers
2) Current newspapers
3) "Gap" newspapers

Newspapers up to 1923 are in the public domain. Newspapers from 1923-1963 either had proactive copyright renewal or went into public domain. Newspapers from 1963-1976 should be checked to see if copyright has been renewed. Those from 1976 on receive automatic renewal. And those from 1998 forward fall under the Sonny Bono Act, which defines the copyright period as "Life plus 70 years." The following URL links to a very useful chart of when works fall into public domain, developed by Lolly Gasaway, University of North Carolina: http://www.unc.edu/%7Eunclng/public-d.htm

Another consideration in copyright is syndicated material, such as newspaper columns. The AP, UPI, USA-Today and others often hold the rights to those materials. To support the effort on currency in South Dakota, some have started to get access through Newsbank. Liz Bishoff offered some options, including offering links only, providing metadata only for copyrighted materials, or not showing articles when the material is copyrighted.

Next Steps

In order to launch the project, people need to know
- Technical Details
- Funding
- Systems

Some of the early steps suggested were to

1) Form a committee to serve as grassroots advocates for the project. This could be expanded by adding other key stakeholders.
2) Develop a list of questions to ask
3) Decide who/how the project will be staffed.
4) Conduct a survey to determine interest in the project, including what types of activities survey respondents would support, and whether or not that support would include financial support.

Liz Bishoff advised going through a basic planning process:

- Define the purpose of the project, including who the audience is.
- Determining what attributes the audience needs and what current systems can do to meet these needs, such as search and retrieval.
- Define the level of comprehensive indexing that is necessary.
- Define the level of accuracy.
- Decide whether the project should use centralized or a distributed system.
- Decide if an open source or a commercial package would be best.
- Develop the project strategy.

In response to a question about the availability of money to hire a leader for the project, Lola Harens (Yankton Community Library) spoke of her experience in writing a National Endowment for the Humanities (NEH) planning grant proposal to launch a program, which in turn led her to receive two subsequent grants. That might be an approach for this project.

Liz Bishoff suggested that the group create both a Request for Information (RFI) and a Planning Project. The RFI could be a way to get people to talk with each other.

Step 1: Greta Chapman proposed that the next step be to bring together all the stakeholders, as an expansion of the leadership team. The State Library and the State Archives are established as leaders. In addition to March 24 meeting participants, others to be invited include the Center for the Book, the South Dakota Endowment for the Humanities; representatives from academic libraries, both regental and public; and the Center for Western Studies. Invitees should be at the decision-making level for their institutions.

Step 2: Add to the Steering Committee/Leadership Group.

Step 3: Develop a meeting to include the expanded group. This meeting should address:

- Project background
- Project purpose
- Audience
- Strategic planning
- Goals and objectives
- Benefits
- Demonstrations of existing projects
- Discussion of how to sell and market the project

- Small group discussion to address
    - Publicity, including developing a PowerPoint presentation to take around the state as a demo, a project brochure, and a project website
    - Fundraising
    - RFI/Technology issues
- Scope (date range for the project)
- Continuing communications

The next meeting will include time for socializing during dinner, and will be held in early to mid-May. Dorothy Liegl and Beverly Lewis will work on creating the expanded group, to include invitees from institutions representing the humanities, academics, genealogy, and newspapers. Greta Chapman, Dorothy Liegl, and Colleen Smith will set the agenda for the meeting.

The Colorado Historical Society was concerned about losing revenue from a possible reduction in microfilm sales; so is South Dakota. It would be beneficial for the State Library, State Archives, and the State Historical Society to meet and create a partnership, and discuss this concern and other ways the institutions can work together on the project.

The project leadership will occur on three levels – a steering committee, the planning committee, which adds representatives from the humanities, academic institutions, and genealogy, and then finally, the stakeholder level.